CH5350: Applied Time-Series Analysis

Models for Linear Stationary Processes

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Classical Time-Series Model

Classical approaches in the early days rested on the philosophy that a series is made up of three components

$$\text{Time-Series} = \text{Trend} + \text{Seasonal Component} + \text{Stationary component}$$

The trend and seasonal components could be combined into a single component under the banner of *deterministic* component.

Classical Approach . . . contd.

Several efficient non-parametric and semi-parametric methods were subsequently developed to realize such a decomposition. The trend usually contains a polynomial type of trend while the seasonal component captures the periodic characteristics, if any.

Extracting the deterministic portions of a series is not trivial, but can be effectively carried out with suitable **regression, smoothing and filtering operations.**

Note: The seasonal component is usually a deterministic periodic signal, and assumed to be uncorrelated with the non-seasonal component.

Modern Approach

In 1970s, a new approach to modelling the seasonal (including the non-stationary and trend components) was introduced.

Unlike the models based on additive approach, **multiplicative models** were postulated. These are more generic in the nature because they take into account the correlation between seasonal and non-seasonal (stationary) components, and also model the integrating (random walk) effects.

The resulting models are known as **seasonal ARIMA (SARIMA) models.**

Models for Stationary Processes

Models for stationary processes

The stationary component, by virtue of definition, cannot be purely explained by a mathematical model but requires the assistance of statistics.

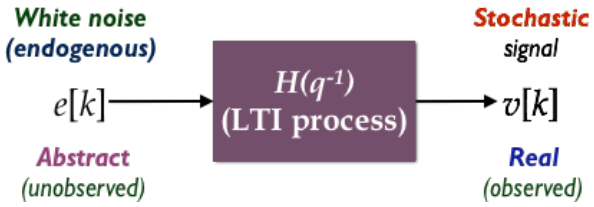
It turns out that a large class of stationary stochastic processes, specifically linear processes, can be explained by mathematical models (convolution / difference equation) driven by forcing functions that are random in nature.

In fact, the forcing function is **the white-noise sequence.**

Key Contributors

1. G.U. Yulé (British, 1871 - 1951)
2. G.T. Walker (British, 1868 - 1958)
3. E. Slutzky (Russian, 1880-1948)
4. H. Wold (Norwegian / Swedish, 1908 - 1992)
5. A. Y. Khinchin (Russian, 1894 - 1959)
6. A. Kolmogorov (Russian, 1903 - 1987)
7. N. Wiener (American, 1894 - 1964)
8. P. Whittle (Newzealander, 1927 -)
9. . . .

Spectral Factorization



The existence of such descriptions (for linear random processes) is centered around a milestone result known as the **spectral factorization theorem**.

The ability to represent $v[k]$ as WN passing through a linear filter, is possible **if and only if the spectral density** (TBD later) of $v[k]$ **satisfies certain mild conditions.**

Absolute convergence of ACVF

While we shall discuss the conditions on spectral density later, it is obvious that the **spectral density function (SDF), denoted by $\gamma(\omega)$ itself should exist in the first place.**

Postponing the formal definition of SDF, at this juncture it is useful to recall from the discussion on non-negative definiteness that the SDF is related to the ACVF through the Fourier transform,

$$\gamma(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \sigma[l] e^{-j\omega l} \tag{1}$$

ACVF should be absolutely convergent

Clearly, for the SDF to exist, and hence the linear representation of the stationary process, the ACVF should be absolutely convergent

$$\boxed{\sum_{l=-\infty}^{\infty} |\sigma[l]| < \infty} \tag{2}$$

Interpretation

For a stationary process to possess a linear filter representation, its ACVF should decay with lag.

- A periodic stationary process, i,e., **a harmonic process,** does not lend itself to a linear representation!

Models and Predictors

Developing a model is as good as developing a predictor. There exists a one-on-one equivalence between a model and a predictor. When the predictor is accompanied by a description of the **prediction error** (what has gone unpredicted), the result is a model.

For stochastic signals, there exists no model that can accurately predict them. However, an optimal model is expected to result in a prediction error that offers no further scope for prediction. Therefore, **the prediction errors associated with an optimal model should possess the properties of a white-noise series.**

Basic idea

Any (stationary) random process can be thought of as consisting of a *predictable* portion plus an *unpredictable* component.

$$v[k] = \hat{v}[k] + e[k] \tag{3}$$

where $\hat{v}[k]$ is the predictable portion and $e[k]$ the unpredictable ideal random process.

- The second term in (3) is an indispensable component of any random process because when it is absent, $v[k]$ condenses to a deterministic process.
- On the other hand, the first term can be absent, in which case, $v[k]$ has white-noise or i.i.d. characteristics.

Conditional expectation: the optimal predictor

Recalling a key result, the best predictor of $v[k]$ is its conditional expectation given its past,

$$\hat{v}^{\star}[k|k-1] = E(v[k]|\{v[k-1], v[k-2], \cdots, v[k-P]\}) \tag{4}$$

- In general, the conditional expectation is **a non-linear function** of the past observations.
- Further, the conditional expectation in (4) is quite difficult to evaluate since the joint p.d.f. of past observations needs to be known.

Linear time-invariant models

If the observations follow a joint Gaussian distribution, the conditional expectation in (4) can be replaced by a **linear model.** (why?)

Replacing the RHS of (4) with a linear function yields,

$$\hat{v}^{\star}[k] = \sum_{i=1}^{P}(-d_i)v[k-i] \tag{5}$$

$$\implies v[k] + \sum_{i=1}^{P} d_i v[k-i] = e[k] \tag{6}$$

where $e[k]$ is the unpredictable random component of $v[k]$.

The model in (6) is known as the **auto-regressive model** of order P.

Remarks

- The negative sign on the coefficients is introduced to have a positive sign on the coefficients of the difference equation for $v[k]$.
- When the true process does not satisfy the (joint) Gaussianity assumption, the linear predictor is sub-optimal, but is at least mathematically tractable and implementable.
- **The difference equation in (5) shares strong similarities with that of a deterministic LTI system - the key difference is that the forcing function is stochastic.**

LTI representation

Equation (6) can be re-written in terms of the transfer function operator.

Example

Suppose that $v[k]$ can be modeled as an AR process of first-order, i.e., by (6) with $P = 1$ (as determined by the sharp cut-off in PACF at lag $l = 1$),

$$v[k] + d_1 v[k-1] = e[k] \tag{7}$$

Bringing in the shift-operator, we can express (7) using the transfer function operator

$$v[k] = H(q^{-1})e[k] \qquad \text{where} \qquad H(q^{-1}) = \frac{1}{1 + d_1 q^{-1}} \tag{8}$$

LTI representations and white-noise

The white noise, thus, acts as a **fictitious input** that drives the random process or the **building block** for time-series models. It is, however, **endogenous**, i.e., intrinsic to the process.

Recall the two other roles of WN in time-series analysis:

- **As a benchmark for testing predictability:** The series (prior to model development) and the residuals (post modelling) are tested for "whiteness."
- **As an integral element of time-series models:** Every random process contains an element of unpredictability, i.e., the WN as its basic element.

Moving Average Models: White-noise as basis

Given P past observations of the process, the modeling objective is to develop a predictor $\hat{v}^\star[k]$ that leaves nothing predictable in the prediction error. Mathematically,

$$\varepsilon[k] = v[k] - \hat{v}^\star[k|k-1] \tag{9}$$

should satisfy (in the **linear** case)

$$E(\varepsilon[k]v[k-i]) = 0, \qquad i = 1, 2, \cdots, \infty \tag{10}$$

- When (10) is applied repeatedly at different time-instants, we are led to a **white-noise** requirement for $\varepsilon[k]$.
- $\hat{v}^\star[k]$ is said to be the **orthogonal projection** of $v[k]$ (in the linear world).

Geometrical interpretation (using Hilbert spaces)

Let $\mathcal{H}$ be the (infinite-dimensional) Hilbert space consisting of all real-valued random variables X with

i. $E(X) = 0$,

ii. $E(X^2) < \infty$

iii. Inner product defined by $\langle X, Y \rangle = E(XY)$.

and all the necessary formalizations).

Let $\mathcal{H}_{k-1}$ be the subspace spanned by $\{v[k-1], v[k-2], \cdots\}$. Denote the orthogonal projection of $v[k]$ on $\mathcal{H}_{k-1}$ by $\hat{v}^{\star}[k]$ and $\varepsilon[k] = v[k] - \hat{v}^{\star}[k]$.

Geometrical interpretation (using Hilbert spaces)

Then,

$$\varepsilon[k] \perp \mathcal{H}_{k-1} \tag{11}$$

In fact, $\varepsilon[k] \perp \mathcal{H}_m, \ \forall m < k$. Also,

$$\mathcal{H}_k = \mathcal{H}_{k-1} \oplus \text{space spanned by } \varepsilon[k] \tag{12}$$

Most importantly (for the discussion), the collection,

$$[\cdots, \varepsilon[k-1], \varepsilon[k], \varepsilon[k+1], \cdots] \tag{13}$$

constitutes an **orthogonal basis** for the Hilbert space spanned by $\{v[k]\}$

Linear Random Processes

Linear Random Process

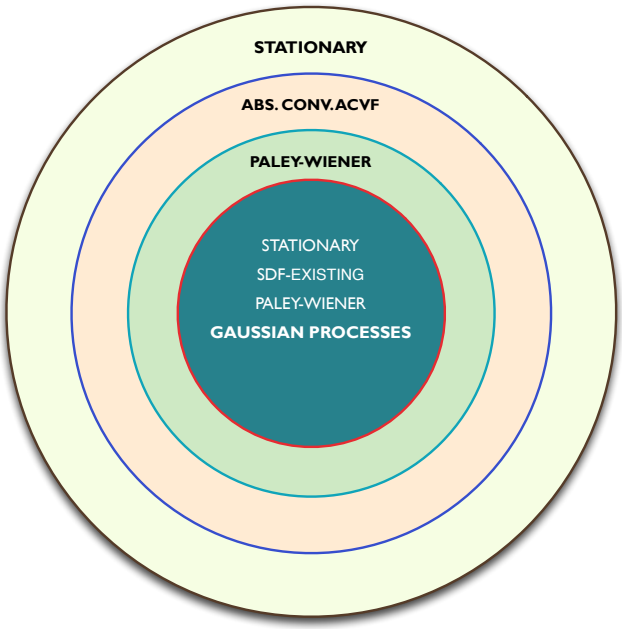
Any stationary process is said to be linear if and only if it can be represented as

$$v[k] = \sum_{n=-\infty}^{\infty} h_n e[k-n] \quad \forall k \quad \text{where} \quad e[k] \sim \text{WN}(0, \sigma_e^2) \tag{14a}$$

$$\text{and} \quad \sum_{n=-\infty}^{\infty} |h_n| < \infty \tag{14b}$$

Implication: Any (weakly) stationary process can be represented as the weighted influences of the past & present ($n \geq 0$) and future ($n < 0$) shock waves.

Linear representations



Linear representations exist only for **stationary, SDF-existent, Paley-Wiener condition satisfying** and yield optimal predictions for **Gaussian** processes.

Remarks

- The condition of absolute convergence (of $\{h_n\}$) is required for the convergence of the infinite sum in (14a) with probability one.
- A weaker requirement is that

$$\sum_n |h_n|^2 < \infty \tag{15}$$

  which guarantees that the sum converges **in the mean square sense.**
- The WN process driving force for the linear process is replaced by an I.I.D. process in certain schools of thought, especially, in non-linear time-series analysis.

Further remarks

- We usually restrict ourselves to causal processes, *i.e.*, series with $n \geq 0$
- With the backshift operator notation, we have

$$\boxed{v[k] = H(q^{-1})e[k]} \quad \text{where} \quad \boxed{H(q^{-1}) = \sum_{n=-\infty}^{\infty} h_n q^{-n}} \tag{16}$$

- The transfer function operator $H(q^{-1}$ can be thought of as a "linear filter", which filters the shock waves to produce the series $x[k]$.

Comparison with convolution form

The model in equation (14) has a striking similarity with the convolution form for deterministic processes. Comparing

$$y[k] = \sum_{n=-\infty}^{n=\infty} g[n]u[k-n] \qquad v[k] = \sum_{n=-\infty}^{\infty} h_n e[k-n]$$

we can draw a few useful analogies

Comparison with LTI deterministic processes

- The coefficient h_n has the some role to play as $g[n]$
- **Stability** of an LTI system requires **absolute convergence of** $g[n]$ **Stationarity** requires **absolute convergence of the coefficients** h_n
- Thus, h_n can be thought of as the **impulse response coefficient** of $H(q^{-1})$

A marked difference between these two forms is that while the input for deterministic systems is known to the user, the input in the stochastic case is fictitious, unknown, random but with known ACF.

Comparison with LTI descriptions . . . contd.

- Just as $G(q^{-1}) = \sum_{n=-\infty}^{\infty} g[n]q^{-n}$ acts as a filter, so does $H(q^{-1}) = \sum_{n=-\infty}^{\infty} h_n q^{-n}$
- It is possible to re-write the infinitely long convolution form for deterministic LTI systems in two different forms:
  - **FIR form** (the impulse response dies down exactly after M instants)
  - **Difference equation** (regressive) form (whenever the IR can be parametrized)
- The **FIR form** is known as the **Moving Average** form (of finite-order) while the **difference equation form** is known as the **Auto-Regressive** form

Modelling viewpoints

1. **Time-series modelling is concerned with the estimation of $H(q^{-1})$ and variance of $e[k]$ given the series $v[k]$**

   - Note that $e[k]$ is fictitious input with specific, but unknown, statistical properties
   - This is one of the prime differences between modelling of linear stationary stochastic and LTI deterministic processes.

Modelling viewpoints . . . contd.

2. The model in (16) is **not unique** in the sense that $\alpha H(q^{-1})$ and σ_e^2/α^2 also offer an equally suitable description of to this process.

   An easy way to resolve this non-uniqueness is by fixing the leading coefficient of the polynomial $H(q^{-1})$ be unity.

$$\boxed{h_0 = 1} \tag{17}$$

3. Even with the fixation of $h_0 = 1$, we are presented with the problem of estimating **infinite unknowns** $h_n,\ n = 1, 2, \cdots$. There are two different ways of overcoming this issue.

Dealing with infinite unknowns: Truncated Model

i. **Assume** that the process is described by a **finite number of IR coefficients**, i.e.,

$$h[n] = \begin{cases} c_n, & 1 \leq n \leq M,\ \ c_n < \infty \\ 0, & n > M \end{cases} \tag{18}$$

leading to the class of **Moving Average** (a.k.a. FIR) models.

Dealing with infinite unknowns: Parametrization of IR

ii. **Parametrize** $h[.]$ in terms of **finite number of unknowns (parameters).** For e.g.,

$$h[n] = \alpha_1 \lambda_1^n \qquad \text{OR} \qquad h[n] = \sum_{i=1}^{P} \alpha_i \lambda_i^P \tag{19}$$

leading to the class of **difference equation form** or the **Auto-Regressive** models.

Moving Average (MA) representations

The moving average representation of order M has the following form

$$\boxed{v[k] = \sum_{n=1}^{M} c_n e[k-n] + e[k]} \tag{20}$$

whose transfer function operator form is

$$\boxed{v[k] = H(q^{-1})e[k], \qquad H(q^{-1}) = 1 + \sum_{n=1}^{M} c_n q^{-n}} \tag{21}$$

In an MA(M) representation, the current "state" of the process is presumed to be **an averaged effect of M past shock waves plus an uncertainty**

Remarks

- From a forecasting perspective, $e[k]$ is the unpredictable part of $v[k]$.
- The **finite-order MA process is always stationary so long as the $\{h[.]\}$ are finite**
- The infinite-order MA process is stationary only when $\{h[n]\}$ is absolutely convergent
- The MA(M) process is also known as M-correlated process.

ACF of MA processes

To determine whether an MA representation is suitable for a given time-series, it is important to study the statistical properties of an MA(M) process and check whether the statistical properties of the given series has a similar structure as well.

The most useful property of interest is the ACF, since it characterizes predictability.

The ACF of finite-order MA process has a special characteristic that not only reveals whether an MA representation is suited for a given series but also facilitates determination of the order of the MA model

Example: ACF of an MA(1) process

Problem: Determine the ACF of a MA(1) process: $v[k] = e[k] + h_1 e[k-1]$

Solution: The solution was derived previously:

$$\rho_{vv}[l] = \begin{cases} 1 & l = 0 \\ \dfrac{h_1}{(1+h_1^2)} & l = \pm 1 \\ 0 & |l| \geq 2 \end{cases} \tag{22}$$

with $\sigma_{vv}[0] = (1 + h_1^2)\sigma_{ee}^2$

The ACF of an MA(1) process abruptly vanishes after $|l| > 1$ (the order of the process)

ACF of an MA(M) process

In general, it is true that the ACF of an MA(M) process dies out abruptly from lag $l = M+1$ onwards. For this reason, the MA(M) process is known as an M-correlated process.

The observation above can be established by working with the **auto-covariance generating function**, which also facilitates easy computation of the ACVF for any MA process given the transfer function operator $H(q^{-1})$

Mapping the ACF of MA model and its parameters

Now we address an important question concerning the estimation of MA models for a given series.

Does there exist a **unique** MA model (for a stationary process) corresponding to a given ACF?

If the answer is in the negative, then further we may have to determine which of the models are admissible.

Uniqueness of MA models for a given ACF

Example

Recall the ACF of an MA(1) process $v[k] = e[k] + c_1 v[k-1]$:

$$\rho_{vv}[l] = \begin{cases} 1 & l = 0 \\ \dfrac{h_1}{(1+c_1^2)} & l = \pm 1 \\ 0 & |l| \geq 2 \end{cases}$$

Given ACF, what is the estimate of c_1?

Example . . . contd.

By solving the quadratic equation using $\rho[1]$ one obtains **two roots that are reciprocals of each other**. By inspection, it is easy to observe that both c_1 and $1/c_1$ produce the same ACF.

Two MA models produce the same ACF

Resolving between the two MA models

Remark

In general, for a given ACF there will always exist two MA models, **one with zeros outside the unit circle and another with zeros inside the unit circle**.

Which model should be chosen?

To answer this question, we turn to **forecasting**, the usual end-use of a model.

The requirement is that the chosen model should produce stable predictions.

Predictions with MA models

Example

Consider the prediction of an MA(1) process: $v[k] = e[k] + c_1 e[k-1]$ given measurements upto k

Denoting the predictions with a hat, we can write

$$\hat{v}[k+1|k] = c_1 e[k]$$

since $e[k+1]$ is unpredictable given any amount of information in the past. The shock wave $e[k]$ can be, however, theoretically recovered from the measurements up to k.

Predictions with MA models

To realize this, observe

$$\begin{aligned} e[k] &= v[k] - c_1 e[k-1] \\ &= v[k] - c_1 v[k-1] + c_1^2 e[k-2] \\ &= v[k] + \sum_{n=1}^{\infty} (-c_1)^n v[k-n] \end{aligned} \tag{23}$$

For the infinite sum on the RHS to converge, $|c_1| < 1$.

Thus, to produce stable predictions, it is important to select the model with $|c_1| < 1$ □

Alternative viewpoint

Observe that we could re-write the MA model for $v[k]$ as a recovery equation for $e[k]$:

$$e[k] + c_1 e[k-1] = v[k] \tag{24}$$

where $v[k]$ is the input to the process and $e[k]$ is the response.

- Notice that the **recovery process is an AR(1) process.**
- We know from prior knowledge that stationarity of an AR(1) process is guaranteed if and only if $|c_1| < 1$.
- Therefore, for stable forecasts, one requires $|c_1| < 1$.

Formal setup

From a formal viewpoint, we are seeking an **inverse model**

$$e[k] = \tilde{H}(q^{-1})v[k] = \sum_{m=0}^{\infty} \tilde{h}[m]v[k-m] \tag{25}$$

such that

$$\tilde{H}(q^{-1})H(q^{-1}) = 1 \tag{26}$$

and $e[k]$ is **stationary.**

Inverse noise model

It can be shown formally that the inverse is (if it is meaningful),

$$\tilde{H}(q^{-1}) = H^{-1}(q^{-1}) \tag{27}$$

i.e., $\tilde{h}[.]$ are the coefficients of series expansion of $H^{-1}(q^{-1})$.

Example

Compute the inverse of the MA(1) model and state the conditions of existence.

$$\begin{aligned} H(q^{-1}) = 1 + c_1 q^{-1} \implies H^{-1}(q^{-1}) &= (1 + c_1 q^{-1})^{-1} \\ &= 1 - c_1 q^{-1} + c_1^2 q^{-2} - c_1^3 q^{-3} + \cdots \end{aligned}$$

Therefore, $\tilde{h}[m] = (-c_1)^m,\ m \geq 0$ and the inversion is meaningful only if $|c_1| < 1$.

Re-writing the prediction equation

For a general linear stationary process, the one-step ahead prediction equation as

$$\hat{v}[k|k-1] = \left(\sum_{n=1}^{\infty} h[n] q^{-n}\right) e[k] \tag{28}$$

$$= (H(q^{-1}) - 1) e[k] \tag{29}$$

$$= (H(q^{-1}) - 1) H^{-1}(q^{-1}) v[k] \tag{30}$$

where we have used the commutativity property of the shift operators. We have thus,

$$\boxed{\hat{v}[k|k-1] = (1 - H^{-1}(q^{-1}))v[k]} \tag{31}$$

Remarks

Equation (31) appears awkward in the sense that $v[k]$ is "apparently" required to predict $v[k]$! However,

Observe that the **leading coefficient in the inverse of noise model is unity. This is always the case** (since we have fixed $h_0 = 1$).

Thus, the first term in the expansion of (31) is no earlier than $v[k-1]$.

Invertibility of MA models

Definition

An MA(M) process is said to be invertible if and only if **all the roots of** $H(z^{-1})$ (which is obtained by replacing q^{-1} by z^{-1}) of the characteristic equation

$$H(z^{-1}) = 1 + \sum_{n=1}^{M} c_n z^{-n} \tag{32}$$

reside outside the unit circle (in terms of z^{-1})

Remark: Naturally, if the roots are computed in terms of the variable z, the roots should reside within the unit circle

Running summary

- Linear random processes can be expressed as infinitely weighted linear combinations of future, present and past shock waves (WN process).
- Descriptions for linear random processes have very strong similarities to that of stable LTI deterministic processes
- Identification (modelling) demands that we place certain restrictions on the TS model (leading coefficient to unity).
- The infinite unknowns are handled by making certain assumptions on the IR coefficients.

Running summary

- The MA(M) model takes birth when it is assumed that $h[n] = 0,\ \forall n > M$.
- MA(M) processes are equivalent to FIR models. They are characterized by ACFs that abruptly vanish at lags exactly after $|l| > M$
- Among the possible MA representations for a process, we admit only invertible MA models
- Invertible models are those whose zeros reside outside the unit circle.

Auto-Regressive Processes

We explore now the other approach to handling infinite unknowns, which **does not subscribe** to the **abrupt drop** of $h[.]$. but rather **parametrizes** $h[.]$ so that it **decays according to some function** (why decay?).

Examples:

i. $h[k] = \alpha_1 \lambda_1^k,\ |\lambda_1| < 1, \alpha \in \mathbb{R}$,

ii. $h[k] = \alpha_1 \lambda_1^k + \alpha_2 \lambda_2^k,\ |\lambda_1|, |\lambda_2| < 1, \alpha_1, \alpha_2 \in \mathbb{R}$.

Now we show that under the above parametrization, the convolution model in (14a) manifests as a **difference equation** model, more popularly known as the **auto-regressive model**.

Parametrization of IR coefficients $\longrightarrow$ DE models

Assume that $h[n] = \alpha_1 \lambda_1^n,\ |\lambda_1| < 1, \alpha \in \mathbb{R}$. Plugging this parametrization into the general linear model of (14a) and writing the equation at two successive instants yields,

$$v[k] = \alpha_1 \sum_{n=0}^{\infty} \lambda_1^n e[k-n]$$

$$v[k-1] = \alpha_1 \sum_{n=0}^{\infty} \lambda_1^n e[k-1-n]$$

A simple algebraic manipulation yields a **difference equation form** for $v[k]$,

$$v[k] - \lambda_1 v[k-1] = \alpha_1 e[k] \tag{33}$$

The AR(1) model

Introducing $d_1 = -\lambda_1$ and **absorbing** α_1 into $e[k]$ (why?), we have from (33), the **first-order** auto-regressive AR(1) model,

$$\boxed{v[k] + d_1 v[k-1] = e[k]}, \qquad |d_1| < 1 \tag{34}$$

- The process is explained purely using its immediate past and hence the name. The uncertainty term $e[k]$ is an integral part of the model.
- It is easy to see that λ_1 is the **root of the characteristic equation** of (34).
- The mapping between the DE form and parametrization is unique.

AR Representations

Extending the foregoing idea to the general parametrization

$$h[k] = \sum_{j=1}^{P} \alpha_j \lambda_j^k, \qquad |\lambda_j| < 1, \alpha_j \in \mathbb{R},\ j = 1, \cdots, P \tag{35}$$

yields **the AR(P) model**

$$\boxed{v[k] + \sum_{j=1}^{P} d_j v[k-j] = e[k]} \tag{36}$$

where λ_j's are the roots (for z) of the characteristic equation of (36) and d_j's are such that these roots are inside the unit circle.

AR(P) models

The transfer function operator form is,

$$v[k] = H(q^{-1})e[k], \qquad \text{where } H(q^{-1}) = \frac{1}{1 + \sum_{j=1}^{P} d_j q^{-j}} \tag{37}$$

In an AR(P) representation, the current "state" of the process is modelled as an averaged effect of its P past "states" plus an uncertainty

- Once again, the parametrization in (35) and the AR(P) model in (36) are uniquely mapped.

Stationary AR processes

The stationarity condition can be derived by using either the equivalence of AR and MA processes or by invoking the analogy of stable deterministic processes.

Stationarity of AR(P) process

A general AR(P) process (see Shumway and Stoffer (2000) for proof),

$$v[k] + \sum_{j=1}^{P} d_j v[k-j] = e[k]$$

is causal & stationary if and only if the polynomial

$$D(z^{-1}) = 1 + \sum_{j=1}^{P} d_j z^{-j} = \Pi_{j=1}^{P}(1 - \lambda_j z^{-1}) = 0 \tag{38}$$

has all roots $\{1/\lambda_j\}_{j=1}^{P}$ (i.e.,in terms of z^{-1}) **outside the unit circle.**

Stationarity of AR processes

- It is important to remember that the constraint is NOT on the coefficients, but on the roots of the polynomial $D(z^{-1})$.
  - If the roots are computed in terms of z, then the requirement is that **all the roots should reside within the unit circle.**
  - The condition is identical to that of a stable LTI deterministic process described by difference equation coefficients $\{d_j\}$
- An alternative definition of a causal stationary AR process is that it possesses an equivalent causal invertible MA representation.

Causality & Stationarity

The foregoing result guarantees causal and stationary representations. Causality is an essential clause there since **non-stationary processes can still be given an AR representation such that** (38) **is satisfied, but one which is non-causal**.

Causality & Stationarity . . . contd.

Example: Explosive AR(1) process

Consider an AR(1) process, with $|d_1| > 1$. Clearly the series is explosive ($v[k]$ grows unbounded). However, we could re-write the series as follows:

$$\begin{aligned} v[k-1] &= (-d_1)^{-1} v[k] + d_1^{-1} e[k] \\ &= (-d_1)^{-2} v[k+1] - (-d_1)^{-2} e[k+1] - (-d_1)^{-1} e[k] \\ \implies v[k] &= \sum_{j=-\infty}^{-1} (-d_1)^j e[k-j] = \sum_{j=1}^{\infty} (-1/d_1)^j e[k+j] \end{aligned}$$

Thus, **an explosive series can be still given a linear stationary representation** since the coefficients of the series are absolutely convergent!

Causality & Stationarity . . . contd.

However, the series is **non-causal** (future shock waves affect the process at present)

In other words, looking at an explosive series "backwards" in time can make it look stationary. However, since non-causal models are of little use in forecasting (non-causal models are used in smoothing time-series), we exclude those AR models that give rise to explosive series and hereafter, understand **causal models as stationary.**

ACF of AR processes

The suitability of an AR model for a given linear stationary process is determined by the ACF signature of the process.

Example: ACF of an AR(1) process

The AR(1) process $v[k] = -d_1 v[k-1] + e[k], |d_1| < 1$ has the ACF

$$\begin{aligned} \sigma_{vv}[0] &= \frac{\sigma_e^2}{1-d_1^2} \\ \rho_{vv}[l] &= (-d_1)^{|l|} \ \forall \ |l| \geq 1 \end{aligned}$$

Remark: In contrast to an MA(1) process, the ACF of an AR(1) process has an **exponential decay.**

ACF of a general AR(P) process

For a general AR(P) process,

$$v[k] = \sum_{j=1}^{P} (-d_j) v[k-j] + e[k] \tag{39}$$

the ACF can be derived by first invoking the definition

$$\sigma_{vv}[l] = E(v[k]v[k-l]) = \sum_{j=1}^{P} -d_j \sigma_{vv}[l-j] + \sigma_{ev}[l], \quad \text{with } \sigma_{ev}[l] = \begin{cases} \sigma_e^2 & l = 0 \\ 0 & l > 0 \end{cases}$$

The ACVF of an AR process satisfies the same difference equation as the original process

General solution to the ACVF of an AR(P) process

Then, writing the ACVF equations separately at lags $l = 0$ and $l > 0$ produces

$$\sigma_{vv}[0] + \sum_{j=1}^{P} d_j \sigma_{vv}[j] = \sigma_e^2 \tag{40}$$

$$\sigma_{vv}[l] + \sum_{j=1}^{P} d_j \sigma_{vv}[l-j] = 0, \; l > 0 \tag{41}$$

The general solution to the ACVF is the solution to the difference equation above

$$\boxed{\sigma_{vv}[l] = \beta_1 \lambda_1^l + \beta_2 \lambda_2^l + \cdots + \beta_P \lambda_P^l \;\; \forall l > 0,} \tag{42}$$

where $\{\lambda_j^{-1}\}_{j=1}^{P}$ are the roots of the CE of (39) and the constants are obtained by solving (40) and (41).

Yule-Walker equations

Equations (40) - (41) are known as **Yule-Walker equations** named after Yule and Walker who discovered these relationships.

AR(2) process

When $P = 2$, the equations are (dropping the subscripts on σ)

$$\begin{aligned} \sigma[0] + d_1\sigma[1] + d_2\sigma[2] &= \sigma_e^2 \\ \sigma[1] + d_1\sigma[0] + d_2\sigma[1] &= 0 \\ \sigma[2] + d_1\sigma[1] + d_2\sigma[0] &= 0 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & d_1 & d_2 \\ d_1 & (1+d_2) & 0 \\ d_2 & d_1 & 1 \end{bmatrix} \begin{bmatrix} \sigma[0] \\ \sigma[1] \\ \sigma[2] \end{bmatrix} = \begin{bmatrix} \sigma_e^2 \\ 0 \\ 0 \end{bmatrix}$$

Example: Y-W equations . . . contd.

The solution is given by

$$\sigma[0] = \frac{\sigma_e^2}{1 - d_1^2 - d_2^2}; \qquad \rho[1] = -\frac{d_1}{1 + d_2}; \qquad \rho[2] = -d_2 + \frac{d_1^2}{1 + d_2}$$

Observe that we could use the Y-W equations to **also estimate the model,** i.e., the coefficients and σ_e^2 given the ACVFs. This is the basic idea underlying a popular method for the estimation of AR models.

Yule-Walker equations

Estimating the parameters of an AR(2) process

When $P = 2$, with the parameters treated as unknowns, the equations are

$$\begin{bmatrix} 1 & \rho[1] \\ \rho[1] & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} -\rho[1] \\ -\rho[2] \end{bmatrix} \Longrightarrow \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho[1] \\ \rho[1] & 1 \end{bmatrix}^{-1} \begin{bmatrix} -\rho[1] \\ -\rho[2] \end{bmatrix}$$

- **Observe that we have used only two of the three Y-W equations to estimate the coefficients** w/o the knowledge of σ_e^2.
- The other equation is useful in computing the variance of the driving force σ_e^2.

Yule-Walker equations for an AR(P) process

The Y-W equations and the theoretical estimates of the model coefficients for a generic stationary AR(P) process are given by:

$$\begin{bmatrix} \sigma[0] & \sigma[1] & \cdots & \sigma[P-1] \\ \sigma[1] & \sigma[0] & \cdots & \sigma[P-2] \\ \vdots & \vdots & \cdots & \vdots \\ \sigma[P-1] & \sigma[P-2] & \cdots & \sigma[0] \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_P \end{bmatrix} = \begin{bmatrix} \sigma[1] \\ \sigma[2] \\ \vdots \\ \sigma[P] \end{bmatrix}$$

$$\implies \Sigma_P \mathbf{d} = -\boldsymbol{\sigma}_P$$

Y-W method for estimating an AR(P) model

It follows that

$$\boxed{\mathbf{d} = -\Sigma_P^{-1} \boldsymbol{\sigma}_P} \tag{43}$$

In practice, the theoretical values are replaced with their respective estimates, resulting in the **Yule-Walker estimates** of the coefficients.

$$\boxed{\hat{\mathbf{d}} = -\hat{\Sigma}_P^{-1} \hat{\boldsymbol{\sigma}}_P} \tag{44}$$

We shall study this method in greater detail in the context of model estimation.

Exponential decay of ACF of AR processes

Recall that unlike for MA processes, **ACF does not provide a distinct signature for determining the order of an AR process.**

- Propagative effects present in AR processes prevent the ACF from detecting the "direct" correlation.
- Any stationary AR process can be re-written as an invertible infinite-order MA process, thereby resulting in an ACF that theoretically vanishes at infinity lag.

Determining the order of AR processes

Two possible solutions:

A. Fit AR models of successively increasing orders (guesses) until we hit upon the "true" order.

B. Discount for the propagative effects to uncover the true order.

It turns out **both these options are identical to using PACF for determining the order.**

Successively increasing orders method

1. Fit AR(p) models of successively increasing orders $p = 1, \cdots, P_{\text{max}}$ (user-defined),

$$v[k] = \sum_{j=1}^{p} \phi_{pj} v[k-j] + e[k] \tag{45}$$

2. Collect the last coefficient $\phi_{pp}, \ ; j = 1, \cdots, P_0$ of each model that is fit.
3. The value of p after which ϕ_{pp} persistently remains at zero is the true order of the AR process.

Order determination . . . contd.

Underlying philosophy: When an AR(p) model is fit to an AR(P_0) process,

$$\phi_{pp} = 0 \qquad \forall p > P_0 \tag{46}$$

The AR models of successively increasing orders can be estimated using Y-W equations (or a least squares method) in a computationally efficient manner using the Durbin-Levinson's algorithm:

Durbin-Levinson's algorithm

1. Fit the best AR model of first order. The optimal estimate is:

$$\phi_{11} = \rho[1]$$

2. The coefficients of AR(p) models of orders $p \geq 2$ can be recursively computed as

$$\phi_{p+1,p+1} = \frac{\rho[p+1] - \sum_{j=1}^{p} \phi_{pj}\rho[p+1-j]}{1 - \sum_{j=1}^{p} \phi_{pj}\rho[j]}$$

$$\phi_{p+1,j} = \phi_{pj} - \phi_{p+1,p+1}\phi_{p,p-j+1} \quad j = 1, \cdots, p$$

Connections with PACF

Comparing the method of computing PACF with the method of using models of successively increasing orders, with minimal effort it is easy to show that the PACF coefficient at lag $l = p$ is the last coefficient of the AR(p) model

PACF coefficient at lag p is ϕ_{pp} :

Thus, one can use the Durbin-Levinson's algorithm to recursively compute PACF coefficients of different lags, as well as to estimate the model coefficients.

Sample PACF calculations using D-L algorithm

PACF computation

Problem: Compute the PACF of a stationary process at lags $l = 1, 2$

Solution: The PACF at lag $l = 1$ is the coefficient of the AR(1) model

$$\phi_{11} = \rho[1]$$

At lag $l = 2$, using the D-L algorithm,

$$\phi_{22} = \frac{\rho[2] - \phi_{11}\rho[1]}{1 - \phi_{11}\rho[1]} = \frac{\rho[2] - \rho[1]^2}{1 - \rho[1]^2}$$

Observe that for an AR(1) process, $\phi_{22} = 0$

This procedure can be continued to compute PACF at higher lags.

Equivalence between AR and MA representations

An invertible MA process possesses an equivalent stationary AR representation and vice versa - a stationary AR process possesses an invertible MA representation.

Example: AR and MA representations

Problem: Find a stationary AR representation of an invertible MA(1) process:
$v[k] = e[k] + 0.6e[k-1]$

Solution: The AR representation can be found by either recursively substituting for past values of $e[k]$ in terms of $v[k]$ or by a long division of the MA polynomial (provided the inverse converges).
We shall use the latter method. Existence of inverse is guaranteed by the invertibility of the MA model.

Example: AR and MA representations

$$v[k] = (1 + 0.6q^{-1})e[k]$$
$$\implies (1 + 0.6q^{-1})^{-1}v[k] = e[k]$$
$$(1 - 0.6q^{-1} + 0.36q^{-2} - ...)v[k] = e[k]$$

$$\therefore v[k] = \sum_{j=1}^{\infty}(-0.6)^j v[k-j] + e[k]$$

Thus, the equivalent AR process is theoretically an infinite-order

Interesting and important observation

Parametrizing the IR coefficients $\{h[.]\}$ of the linear random process in (14a) is equivalent to **parametrizing the ACVF** of the stationary process.

Parametrizing ACVF in turn amounts to **parametrizing the spectral density function** of the process!

ARMA representations

In reality, several processes have **mixed effects,** i.e., both MA and AR effects, or it is that we require a fusion of both to explain the process in a parsimonious manner:

For this purpose, we turn to a more general representation that contains mixed effects. This is the auto-regressive moving-average (ARMA(P, M)) representation

$$v[k] = H(q^{-1})e[k] \tag{47a}$$

$$H(q^{-1}) = \frac{C(q^{-1})}{D(q^{-1})} = \frac{1 + \sum_{i=1}^{M} c_i q^{-i}}{1 + \sum_{j=1}^{P} d_j q^{-j}} \tag{47b}$$

On ARMA representations

ARMA models that are **both invertible and stationary** are allowable representations for causal, stationary processes

- ARMA processes simplify to AR or MA models depending on whether $C(q^{-1}) = 1$ and $D(q^{-1}) = 1$ respectively
- An ARMA process can always be given an equivalent AR or MA representation. In either case, the resulting models are of infinite-order

ACF of an ARMA process

There are primarily two different ways in which one can calculate the theoretical ACF of a ARMA(P, M) process:

A. Convert the given ARMA process to a MA process and use existing results
   - Write the difference equation in MA series (will be of infinite order)

B. Write the difference equations for the ACF.
   - The first K set of equations for lags $l = 0, \cdots, K-1$ where $K = \max(P, M+1)$ are linear in the ACVFs. They have to be solved simultaneously
   - The second set of equations are the recursive relations like those that arise in AR process. These are solved recursively for lags $l = K, K+1, \cdots$

On the order determination of ARMA processes

Neither the ACF nor PACF can provide a practically useful distinct signature to determine the order of the numerator and denominator polynomials

---

In practice, one uses information-theoretic measures such as Akaike's Information Criterion **(AIC)**, Bayesian (Schwartz) Information Criterion **(BIC)** or a machine-learning metric such as Minimum Description Length **(MDL)** to determine the appropriate order.

Choice of representation

A commonly encountered question is: which form to choose? - MA or AR or ARMA?

- There is no definitive answer since all forms are inter-convertible. A few guidelines, however, exist primarily from an estimation point of view

Which model?

- **No model is a true description of the process!**
- A model only provides an approximate abstraction of the process evolution with time (or space, frequency, etc.)
- Keep the model as simple as and whenever possible (principle of parsimony)

  Choose that model description which contains few parameters without compromising significantly on the quality and predictive ability of the model
- Over-parametrized model implies more parameters to estimate and hence larger error $\Longrightarrow$ lower reliability of the model (we give a theoretical proof later)

Commands in R

Commands	Functionality
`ts, as.ts`	Create time-series objects
`ARMAacf`	Theoretical ACF and PACF for a given ARMA model
`acf2AR`	Derive an AR model for a given ACF
`arima.sim`	Simulate AR(I)MA models
`acf, pacf`	ACF and PACF estimates with plots
`arima, ar`	Estimate AR(I)MA and AR models, respectively